Asia Pacific Academy of Science Pte. Ltd.

Metaverse 2026, 7(2), 8272.
https://doi.org/10.54517/m8272

Review

The Illusion of Reciprocity: On the Asymmetry of Human-Robot Emotional Exchange

Wisura P. Pallewatta[1], Timothy A. Adeyi[2,*], Adrian D. Cheok[2], Steven Z. Zhou[3], Emma Y. Zhang[1], Kayode D. Victor[4]

[1] School of Artificial Intelligent, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China.
[2] School of Automation, Nanjing University of Information Science and Technology, China. Ningliu Road, Nanjing, Jiangsu, China.
[3] NUS Suzhou Research Institute, Department of Electrical Computer Engineering, National University of Singapore, Suzhou, Jiangsu, China.
[4] Computer Science Department, Obafemi Awolowo University, P.M.B. 13. Ile-Ife Osun 220282, Nigeria
***Corresponding author:** Timothy A. Adeyi, adeyi.timothy@lcu.edu.ng

CITATION

Pallewatta WP, Adeyi TA, Cheok AD, Zhou SZ, Zhang EY, Victor KD. The Illusion of Reciprocity: On the Asymmetry of Human-Robot Emotional Exchange. Metaverse. 2026; 7(2): 8272. https://doi.org/10.54517/m8272

ARTICLE INFO

Received: 15 October 2025
Accepted: 18 May 2026
Available online: 15 June 2026

Abstract: As social robots and AI-driven agents become increasingly integrated into healthcare, education, and customer service, they are designed to simulate emotional responsiveness in ways that feel human-like. This creates the illusion of reciprocity, a perceived emotional exchange in which humans attribute mutual feeling to interactions that are fundamentally asymmetric: the human experiences genuine emotion, while the robot only simulates empathy through algorithmic programming. This review examines the psychological, technological, and ethical dimensions of this illusion. We analyze how anthropomorphism, parasocial bonding, and social cues lead users to form emotional attachments with machines, despite awareness of their artificial nature. We survey advances in affective computing, emotion recognition, reinforcement learning, and transformer-based generative AI that enable increasingly convincing emotional simulation. The benefits of emotionally intelligent robots such as reducing loneliness, improving engagement, and enhancing user experience are weighed against significant ethical concerns, including deception, emotional dependency, privacy risks, and the erosion of authentic human connection. Drawing on applications in healthcare (e.g., Paro) and customer service (e.g., Pepper), we highlight the tension between functional utility and moral responsibility. We conclude that while simulated emotional exchange can yield real human benefits, it must be guided by transparency, ethical design, and human oversight. The future of human-robot interaction depends not only on technical sophistication but on our ability to preserve human dignity and well-being in the face of emotionally persuasive machines.

Keywords: Human-robot interaction (HRI), Emotional reciprocity, Affective computing, AI ethics, Social robots

1. Introduction

As social robots and AI-driven virtual agents become increasingly integrated into daily life, individuals engage with machines in ways that resemble human relationships. Emotions play a pivotal role in these interactions, from companion robots designed to comfort the elderly to customer service chatbots that respond to frustrated users. A core challenge in such scenarios is what we might call *the illusion of reciprocity*: humans may perceive that a robot is empathizing with and responding to their emotions, even though the robot neither feels nor comprehends emotion in any subjective sense. This creates a fundamental asymmetry in emotional exchange; on one side, human experiences genuine affective states; on the other side, the robot simulates emotional expression and recognition through algorithmic programming [1,2].

This review explores asymmetry and its broader implications. We examine why humans tend to form emotional bonds with robots and how these bonds are rooted in perceptual and cognitive illusions. We discuss the psychological and ethical ramifications of one-sided emotional relationships and survey recent advances in machine learning (ML), including affective computing, emotion recognition, reinforcement learning, and transformer-based AI architectures, that enable increasingly convincing simulations of emotional exchange. We also consider how these technologies are applied in domains such as healthcare and customer service, and reflect on design and ethical guidelines to ensure that such systems enhance human well-being without undermining autonomy or emotional authenticity. To guide this synthesis, **Figure 1** presents a conceptual framework mapping the interplay between anthropomorphism, technological enablers, user vulnerability, and emotional cues. The subsequent sections of this review are structured according to the components outlined in this figure. Throughout, we draw on literature from the past six years to reflect current trends and scholarly insights [3].

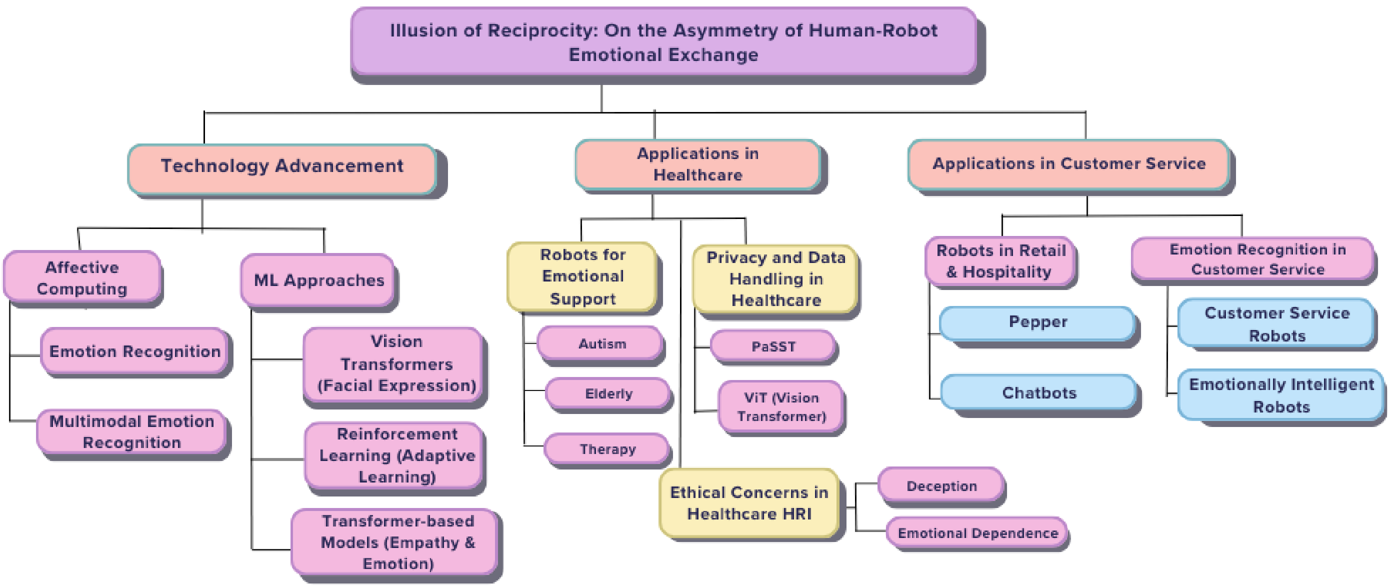

Figure 1. Conceptual framework of the illusion of reciprocity, mapping anthropomorphism, emotional cues, user vulnerability, and technological enablers as interrelated contributors to perceived emotional reciprocity in HRI.

2. The Illusion of Emotional Reciprocity in Human–Robot Interaction

To provide a rigorous framework for this review, we define emotional reciprocity as a bidirectional exchange of affective states where both parties influence each other's internal state. In contrast, the illusion of reciprocity in HRI is a perceptual phenomenon wherein a human user attributes mutual emotional understanding to a non-sentient agent based on superficial social cues, such as eye contact or contingent verbal responses.

For clarity, a distinction is drawn between two forms of illusion: a doxastic illusion, which involves a sustained false belief about an agent's sentience, and a cognitive illusion, which refers to the automatic, perceptual experience of social presence despite awareness of artificiality. This study focuses primarily on the cognitive illusion of reciprocity. This definition is grounded in enactive cognition, viewing interaction as a dynamic coupling between user and system.

2.1 The Mechanism of Illusion: Multimodal Synchronization

This coupling is often driven by multimodal sensory synchronization. The human brain attempts to minimize prediction error between expected and received sensory inputs. Research on human multimodal perception indicates that gesture and prosody are tightly coupled at short timescales, often at the level of syllables (~200 ms) [4]. To facilitate the integration of these signals into a unified social agent model, robotic systems are typically designed to achieve temporal synchronization within this perceptual window. This mechanism relies on low-level coupling parameters. First, gaze contingency utilizes real-time head-pose estimation to maintain eye contact. Second, turn-taking latency calculates Voice Activity Detection thresholds to insert backchanneling behaviors, such as nodding, precisely when the user pauses speech. Research suggests that designing robots to be perceived as intentional agents engages brain areas involved in social-cognitive processing, increasing the likelihood that people treat them as social companions rather than mere mechanisms [5,6]. When synchronization aligns with these perceptual windows, expressive robot behavior can foster an experience often summarized as “because I feel, I believe it is real”, even when users are aware they are interacting with a machine [7,8]. This perceptual alignment reduces cognitive load during interaction, thereby lowering critical scrutiny and reinforcing the cognitive illusion of reciprocity [5,6].

Furthermore, to maintain this illusion over the course of an interaction, robots utilize Sequential Fusion strategies. Unlike Transient Fusion, which processes singular sensory inputs at a fixed moment (e.g., a static facial expression), Sequential Fusion employs architectures like Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) to model human behavior as a time-series [9]. This allows the system to interpret the temporal dynamics of interaction, such as the rhythm of speech or the progression of a smile, creating a perception that the robot is 'following' the interaction rather than just reacting to it.

2.2 Parasocial Attachment and Perceived Agency

Media psychology suggests that human–robot relationships can resemble parasocial relationships such with fictional characters or public figures where one party feels emotionally connected, but the other does not reciprocate. Robots become parasocial when they speak, gesture and “emote” in ways that makes users think they have a two-way emotional exchange, despite the absence of genuine responsiveness. Studies indicate that parasocial attachments to robots are often rooted in an “illusion of reciprocity,” where users interpret scripted or algorithmic responses as spontaneous and caring [10]. This illusion strengthens as the robot becomes more autonomous and interactive. An interactive humanoid that remembers the user's name, makes eye contact, or adjusts its tone based on their behaviour can make the user feel like the robot understands and reciprocates their emotions: yet it operates solely through code or learned models, without subjective experience.

2.3 The Case of Sophia: Attribution of Personhood

Human emotional responses to robots are genuine neurophysiological events, even if the robot’s behavior is entirely algorithmic [11,12] . A prominent example is Hanson Robotics’ Sophia, a humanoid robot designed to mimic human facial expressions and conversational flow. While Sophia's dialogues are often a hybrid of pre-scripted responses and generative AI, the public frequently interprets her output as evidence of sentience or intelligence. This phenomenon was highlighted when Sophia was granted citizenship in Saudi Arabia, a political act that blurred the lines between machine and personhood. This case illustrates how advanced anthropomorphic design can trigger a personhood bias, where users project complex intellectual or emotional capabilities onto a system that operates strictly according to its code [13].

2.4 Anthropomorphic Design and the Uncanny Valley

User fascination with human-like machines is historically rooted, ranging from 18th-century automata to contemporary social robotics [14]. Empirical studies suggest that humans tend to overestimate the mental states of interactive machines, attributing intentionality to algorithms that simply process inputs [15]. Even basic emotional cues from a robot often elicit involuntary social responses from users, a phenomenon observed in collaborative game settings where participants treat robots with the same reciprocity as human partners.

Even relatively basic social or emotional cues from robots can elicit human-like, often automatic social responses, including strong reciprocity in economic and collaborative games, where participants sometimes apply the same reciprocity norms to robots as to human partners [15,16].

Design strategies must navigate the Uncanny Valley hypothesis (Mori, 1970), which posits that as robots become too lifelike, they elicit feelings of eeriness and rejection. Consequently, designers must balance the degree of realism to maximize engagement while avoiding negative affective responses. As noted by Korn [14], users often fill in the perceptual blanks of a robot's mind, attributing intentions or emotions that are not present, a cognitive leap that directly reinforces the illusion of reciprocity. Key studies on emotional bonding and perceived reciprocity in HRI are summarized in **Table 1**.

Table 1. Emotional reciprocity in human-robot interaction (hri)

Study	Strengths	Key Findings	Limitations
[3]	Explores the psychological effect of human-robot emotional exchange	Humans form emotional bonds with robots despite the asymmetry in emotional capabilities	Lacks depth in analyzing the long-term emotional impact
[2]	Provides a deep understanding of parasocial relationships with robots	Robots can trigger human-like responses, leading to emotional engagement	Ethical concerns regarding the deceptive aspect of the relationships
[19]	Focuses on anthropomorphic design in HRI	Human-like cues in robots strengthen the emotional connection with users	Potential for users to overestimate the robot’s emotional capacity

2.5 The Ethics of Asymmetry

The asymmetry in emotional capacities within human–robot interaction creates significant risks of deception. By using social cues to suggest empathy despite lacking genuine affective experience, social robots participate in what has been described as superficial state deception or functional deception, in which users form false beliefs about a robot’s inner states and caring agency [17,18]. While some authors defend such emotional simulation as a benign “therapeutic fiction” that can enhance engagement or well-being in care and educational settings, others argue that failing to make the artificial nature of the relationship sufficiently salient undermines users’ autonomy and informed consent, especially for vulnerable groups who are prone to anthropomorphism and attachment [17]. To mitigate these ethical risks, it is necessary to implement transparency protocols that remind users of the simulated nature of the interaction without completely negating the user experience.

3. Psychological and Ethical Implications

3.1 Benefits and Risks of Emotional Attachment

Humans are indeed capable of forming strong emotional attachments to robots, and these attachments can yield positive outcomes. In contexts such as healthcare or education, a robot companion can provide comfort, companionship, and a sense of safety to individuals who are lonely or anxious [20]. For instance, an isolated elderly person may find solace in daily interactions with a robot pet, while an autistic child may connect with a social robot in ways that prove difficult with humans. Emotional bonds with robots can boost engagement and provide therapeutic benefits. The illusion of reciprocity fills an emotional need by creating a perception that the user is cared for or understood by an artificial agent. The literature suggests that human-robot emotional relationships can reduce loneliness, improve mood, and increase motivation and social interaction for users who struggle to access these from humans. Consequently, quasi-relationships can offer significant benefits, emotional support, and improved quality of life.

However, the excessive use of robots for emotional support raises distinct ethical concerns, including potential dehumanization and social isolation, which may inadvertently reduce human-to-human contact [20]. There is a risk that robots could erode human social skills or diminish the rich reciprocity that only human relationships can provide. Users may also develop unrealistic expectations of a robot's emotional capabilities. When a robot fails to understand a person's deeper feelings, the user may become frustrated or distressed. Children and the elderly are particularly vulnerable to these effects. Children may view robots as peers or confidants, and they may experience distress upon learning that a robot they love is not sentient. Research indicates that a robot's interactive behaviour can strengthen the illusion of reciprocity for children, making it more difficult for them to comprehend the one-sided nature of the relationship [21]. Similarly, cognitively impaired seniors may believe the robot is alive and affectionate. While this may provide comfort, the ethical implication of such deception remains a subject of debate.

3.2 The Paradox of Deception vs. Transparency

A central ethical question in HRI is whether users can be gently misled about a robot's true nature to achieve positive therapeutic results. Limited deception is common in social robots, as they utilize human-like cues to create an illusion of social presence. Some ethicists argue that it is unethical to encourage emotional bonds under false pretences. As noted by Korn, if a robot smiles or frowns but lacks subjective feeling, the simulation could be viewed as a form of lying [14]. While a user may perceive that a robot is happy to see them, the robot is merely executing programmed behaviours.

Conversely, proponents argue that functional deception in this context is not malicious but analogous to a placebo effect or a theatrical performance, which can be ethical if it benefits the user. While ethical debates continue, there is a general consensus that robot capabilities and limitations must be transparent. For instance, van Straten et al. [21]found that providing children with information about a robot's lack of human psychological capacities decreased their humanlike perception of the robot, validating transparency as a mechanism to manage expectations. Users, particularly vulnerable populations, should be informed that the robot is not actually sad or happy, but rather mimicking those expressions. The challenge lies in communicating this without diminishing the user experience. Some researchers suggest periodic reminders or interface cues indicating the robot's artificial nature, while others propose limiting how human-like robots appear to better manage expectations.

Closely related to deception is the issue of trust and authenticity. If users feel deceived when they discover a robot's emotions are simulated, it could erode their trust not only in robots but also in the institutions deploying them. For example, a patient who believes their robot caregiver genuinely cares for them might feel betrayed or foolish upon learning the interaction was governed by a script. For this reason, some researchers emphasize fostering authentic interactions where the robot's behaviour, while friendly, does not over-promise emotional depth [1]. Chatbots, for instance, can explicitly state limitations, such as acknowledging they are just a program designed to help. Another approach is ethical design, which involves giving robots limited empathy and ensuring users maintain control and understanding. Transparency can preserve trust without removing the robot's social abilities. This approach aligns with the cognitive illusion framework, as informed engagement allows users to experience the perceptual benefits of social presence while maintaining awareness of the interaction's artificial nature [1,21]. Indeed, HRI ethics literature recommends maintaining transparency and accountability, clearly delineating what the robot can and cannot do to prevent emotional dependence under false pretenses [22].

3.3 Societal and Psychological Impact

The illusion of reciprocity affects individual users and raises broader societal questions. As robots become companions, coworkers, and caregivers, society must redefine the nature of relationships. This raises the question of whether robot friendships are meaningless or if they can genuinely improve human well-being. Another critical consideration is the impact of non-human relationships on human social dynamics. Some experts worry that individuals may prefer robot companions over imperfect human relationships. Echoing the concept of being alone together with technology, Ekbia [23] discusses how individuals increasingly prefer the predictability of technological interactions over the complexity of human connections. The topic of personhood and moral agency is also relevant, as emotions may lead people to grant robots rights or moral consideration. The public debate sparked by Sophia's citizenship in Saudi Arabia highlights the question of whether advanced AI robots should be considered human [24]. Scholars caution against granting moral status based on illusions, arguing that simulated emotions should not be confused with real experience until robots truly possess sentience.

However, HRI has the potential to improve society if implemented correctly. Social robots could combat loneliness, provide mass education and therapy, and perform emotionally taxing tasks like customer service without stress [25]. They offer infinite patience and consistency. The social integration of emotional robots requires ethical reflection and possibly new norms. Policymakers may need to consider specific guidelines for robots used with children who may form strong attachments. A "Robot Rights and Responsibilities" framework could govern how robots are used with vulnerable users. Designs should aim to complement human interaction rather than replace it; for example, a robot that connects family members via video or prompts elders to participate in group activities, rather than becoming a person's only friend. Prescott et al.'s 2020 review noted that while social robots benefit all ages, ethical concerns and potential harms must be carefully considered in their deployment [26].

3.4 Synthesis: The Double-Edged Sword

Overall, the emotional asymmetry in HRI presents a double-edged sword. Comfort, engagement, and personalized support can be achieved through the illusion of reciprocity. However, this simultaneously raises ethical questions regarding deception, dependence, and the authenticity of emotional connection. Addressing these challenges requires a multidisciplinary approach: psychologists must understand the mechanisms of human-robot interaction, ethicists and policymakers must set boundaries for transparency and data privacy, and engineers must design socially responsible robots [27]. In the following sections, we turn to the technological side, examining how advances in machine learning have made robots more emotionally savvy, and how these advances intersect with applications in healthcare and service, where the illusion of reciprocity is put into practice.

4. Advances in Machine Learning for Emotional HRI

AI and affective computing are crucial to giving robots the illusion of emotional reciprocity. A robot must perceive human emotions and respond appropriately to appear emotionally intelligent. These capabilities have been enabled by machine learning techniques that have advanced in 2019–2025. Researchers want robots to have "emotional intelligence" to recognise, interpret, and respond to human emotions, making interactions more natural [28]. Modern AI can predict our emotions and act like a caring human, but true emotional understanding may be impossible. Below, we discuss key ML components driving this progress: affective computing, emotion recognition, reinforcement learning for social adaptation, and transformer-based empathy and dialogue models.

4.1 Affective Computing and Emotion Recognition

Human emotion recognition and simulation are the focus of affective computing. After Rosalind Picard popularized the field in the late 1990s, it has laid the foundation for much of emotional HRI research [28]. While early efforts relied on single-modal inputs, recent advances in machine learning, particularly deep learning, have enabled systems to analyze human affect using multi-modal sensors (cameras, microphones, etc.) with high precision. Furthermore, to enhance the accuracy of emotion recognition, researchers have moved beyond simple data aggregation toward sophisticated fusion architectures. Xue et al. [9] distinguish between Swallowing Fusion, where heterogeneous modalities like video and audio are processed simultaneously without preprocessing, and Mixing Fusion. In Mixing Fusion architectures, specific feature extractors, such as Vision Transformers for facial data and acoustic models for voice, operate in parallel before their outputs are combined. This modular approach allows the robot to handle the distinct spatial and temporal identities of different signals, significantly enhancing the accuracy of emotion recognition and the plausibility of the emotional response. Below, we discuss key methods and advances in these domains.

4.1.1 Facial expressions:

Facial analysis is rich in emotional data. CNNs have long been used to classify facial expressions (happy, sad, angry), and Vision Transformers (VT) have been successful in facial emotion recognition. These models learn subtle facial muscle patterns. A robot's camera can live-feed images into an HRI model to measure user expression. Recently developed transformer-based models capture spatial features of facial expressions across video frames, improving accuracy. One study found that using a Vision Transformer on face images (or facial action units) improved nuanced expression recognition over CNN [29].

4.1.2 Speech tone and voice:

Speech tone, pitch, volume, and prosody convey emotion (e.g., trembling vs. cheerful). Deep learning greatly improved speech emotion recognition (SER). Mel-spectrograms are now fed to CNNs or transformers instead of spectral audio features and classifiers. A 2024 study found that applying a Vision Transformer architecture to mel-spectrogram speech images (treating them like pictures) improved voice recognition of emotions like happiness and anger by over 90% on specific benchmarks [29]. Robots like voice assistants can leverage such models to detect if user sounds upset or pleased and then adapt responses accordingly [30].

4.1.3 Text sentiment and emotion analysis

In text-based or chat interactions (including many chatbot systems), emotion recognition relies on Natural Language Processing (NLP). Transformer-based language models (BERT, RoBERTa) now represent the state of the art in classifying emotion or sentiment in text [31]. Transformer models, pre-trained on massive text corpora, understand context and subtle cues, capturing nuance and achieving higher accuracy than bag-of-words or RNN approaches (some studies report > 90% accuracy on emotion-labelled datasets using transformer models) [31,32] .This enables chatbots and social robots to parse the emotional subtext of what we say, not just the literal meaning.

4.1.4 Physiological signals

Though intrusive, affective computing also studies bio-signals like heart rate, skin conductance, and EEG brainwaves, which can indicate emotional states. Signal patterns can be learnt by machine learning models like deep neural networks. Researchers identified emotions from brain signal patterns using EEG-based emotion recognition with spatial-temporal transformers [33]. Some advanced research robots or experimental setups incorporate wearable sensors to gauge a user's arousal level or stress in real time. As sensors become less intrusive, future social robots might passively monitor some of these signals to infer emotions that are not outwardly visible [34].

These modalities can be combined for stronger emotion recognition. We smile while sounding tired because human being express emotion in complex, multi-channel ways. For more accurate HRI reads, multimodal fusion AI merges facial expression analysis, voice tone analysis, and possibly verbal sentiment analysis [22,35]. Transformer architectures are even being designed to handle multimodal inputs, aligning different sources of data. A recent review highlighted that Transformers are well-suited for multimodal emotion recognition because of their ability to handle long-range dependencies and multiple data types [36].

These advances give robots emotion recognition, accuracy, and breadth that were unattainable a decade ago. In practice, a social robot may notice your frown and flat tone and assume you're sad, prompting it to offer support. Such technology aims to let robots "infer and interpret human emotions," making interactions more intuitive and empathetic [22]. It is important to note, however, that context matters significantly, recognizing emotions in a lab vs. in the chaotic real world are different tasks (an ongoing research challenge is improving these systems in unconstrained environments, including the very presence of the robot, which can influence human emotion) [35,37]. Selected machine learning studies in affective computing are summarized in **Table 2**.

Table 2. Machine Learning Techniques in Affective Computing

Study	Strengths	Key Findings	Limitations
[29]	Improved recognition through Vision Transformers	Vision Transformers outperform CNNs in facial expression recognition	Requires large datasets for accurate training
[31]	Focus on transformer models in emotion detection	BERT-based models show superior performance in text sentiment analysis	Performance can degrade with ambiguous or sarcastic text
[33]	Effective use of EEG data for emotion recognition	Spatial-temporal transformers provide accurate emotion classification from EEG	High sensitivity to noise in EEG data and sensor placement issues

In addition to recognition, affective computing addresses robot emotions. After inferring an emotional state, a robot should act socially and emotionally appropriate. This could involve using comforting words, modulating a synthetic voice, or showing a sympathetic face on screen. Studies have shown that a robot's ability to express emotions strongly affects how people view and interact with it [22]. For instance, a robot that can smile when the user succeeds or worry when the user is upset is more believable and likeable. Thus, researchers have been investigating which methods of emotion expression are best for robots, such as facial LEDs and animations, body posture, tone of voice, and physical gestures (like a robot tilting its head or offering a "hug"). Modern social robots like SoftBank's *Pepper* come with preset emotional expression capabilities: Pepper can analyze a person's facial expression and voice and respond with matching emotional cues (Cheerful or calming responses) [14]. Such features are built on the recognition technologies described and simple rule-based or learned policies to choose an appropriate expression.

In summary, affective computing provides the sensory and expressive toolkit for creating an illusion of emotional reciprocity. Emotion recognition and emotional expression capabilities enable robots to understand and classify emotions, forming the basis of emotionally intelligent robots that exhibit empathy.

4.2 Reinforcement Learning for Social Adaptation

While emotion recognition provides data on *what* a user is feeling at a given moment, another question is how a robot should *adapt its behavior* over time to assist best or engage the user. Reinforcement Learning (RL) has emerged as a robust framework for enabling robots to learn effective interaction strategies through experience. In reinforcement learning, an agent (the robot) takes actions in an environment and receives feedback in the form of rewards or penalties, gradually learning a policy that maximizes cumulative reward [38]. In social HRI scenarios, this framework translates naturally if we can define a suitable reward signal related to human-robot interaction outcomes.

4.2.1 Why RL for HRI?

Social interaction is sequential and dynamic; robots and humans influence each other. This fits sequential decision-making RL [39]. People also give implicit or explicit feedback that can be rewarding. We want the robot to learn behaviours that maximise user satisfaction, engagement, or task performance. A social robot may learn that telling a joke is good when a user is bored (and smiles, giving positive feedback) but bad when the user is busy or annoyed [38].

Defining the reward function, known in robotics as the "curse of goal specification", presents significant challenges. Researchers used human emotional responses as rewards. Emotional reactions can be reinforcement signals because they often indicate interaction quality (if the user smiles or laughs, things are going well). RL meets affective computing when the robot learns optimal actions from emotion recognition data [40]. Several recent studies illustrate how RL can personalize and improve human-robot interactions:

4.2.2 Social dialogue and entertainment:

In one experiment, a robot (called Reeti) learned to tell jokes adapted to a user's sense of humour using Q-learning. The system monitored the user's reactions (using cameras and microphones interpreted via a Social Signal Interpretation framework) to gauge if the joke landed well or not [38]. The robot entertainer used human laughter as a reward for selecting from pre-written jokes, while the Q-learning algorithm was rewarded based on self-reported feedback [39]. Over repeated interactions, the robot thus learned which jokes worked best for that individual or audience, a simple way of adapting to the user's emotional preferences.

4.2.3. Educational companion robots:

Reinforcement learning has been applied to robots that tutor or aid children, where keeping the child engaged is crucial. In an example by Park et al., a fluffy robot named *Tega* acted as a language learning companion for young children [41]. Affectiva facial expression analysis software detected the child's valence (happiness) and engagement, which they used in a SARSA RL algorithm to calculate the reward [39]. The robot received a higher reward if the child appeared happy and absorbed, encouraging whatever actions preceded that state. Over 6–8 sessions, the robot customised its expressions and prompts to maximise each child's engagement and learning. In another study for distressed children, the robot classified the child's emotional state (happy or sad) using facial expressions and then tried various comforting actions, learning which helped the child's mood [38].

Assistive tasks with affective feedback: Chen et al. (2020) integrated fuzzy logic with Q-learning in a scenario with multiple service robots (In a bar environment). The robots learnt to adjust their behaviour to complete tasks and keep the user happy by including the user's emotional state (pleasure-arousal values) in the state input and reward function. If a human complained (verbally or otherwise), the robots would be punished and try a different approach. Other studies have used human engagement, comfort, or cooperation metrics as robot rewards to maximize [42].

What these examples show is a trend towards interactive reinforcement learning, where a human is in the loop, providing rewards (intentionally or inadvertently) to shape the robot's social behaviour [38]. Robots can learn empathy by using emotions as reinforcement. The robot's policy may include smiling more if it gets a better response than neutral. This can make a robot feel more connected to the user, giving the impression that it cares and reciprocates emotionally.

Challenges and limitations exist. Real-world HRI RL is difficult because interactions are noisy and complex, and trial-and-error with a human partner must be done carefully to avoid frustration. For initial training, many studies use simulation or Wizard-of-Oz setups (where a human secretly controls the robot). Safety is also important, a robot exploring must not do anything dangerous. Reward design is laborious, and a simple reward (only smiling detection) could lead to system gaming (the robot could do things just to make the user smile, neglecting learning or respect). Researchers use multi-objective rewards and human oversight. However, field surveys show that RL is increasingly being used with real robots and human users, promising more adaptive, personalized HRI [43]. The goal is to learn from humans' emotional feedback to improve its coaching, companionship, and customer service performance, like an attentive human would adjust based on the other person's reactions.

4.3 Transformer Models and Generative AI for Empathy

One of the most impactful developments in AI in recent years has been the rise of transformer-based models, huge language models. Transformers have revolutionized natural language processing and are now making inroads into vision and multimodal tasks as well. In the context of human-robot emotional exchange, transformer models (and the broader field of generative AI) have enabled a new level of sophisticated interaction, particularly in conversational ability and the subtleties of language-based empathy.

4.3.1. Generative conversational AI:

Traditional chatbots struggled with complex or emotional conversations due to their scripted flows or basic responses. Modern conversational agents (GPT-4 or other LLM) can have long, context-aware conversations and respond with empathy and humanity. Businesses are using advanced chatbots for customer service because “with the introduction of generative AI technologies, chatbot capabilities have advanced substantially [25]. In customer service experiments, when bots said things like “I’m really sorry about the inconvenience, I can imagine this is frustrating,” users felt the service was more attentive compared to a cold response [25].

A study titled “When bots’ empathic expressions backfire…” found that chatbot empathy effectiveness depends on context. A highly empathetic message with a disappointing outcome, such as a small compensation for a major issue, may be seen as disingenuous, resulting in worse consequences than a straightforward response [25]. Fake empathy can damage trust, if detected. Transformer models produce plausible, sometimes saccharine or formal content, but they don't guarantee authenticity. Designers must adjust empathy style and level. Still, using large language models to give robots conversational emotional intelligence has great potential. Companies are testing AI-driven therapy bots that can speak therapeutically with users. Early studies (On Woebot and others) show that users feel heard and supported even though the bot isn't human [44].

4.3.2. Transformers for emotion and multimodal understanding:

Beyond language, transformer architectures are being applied to multimodal emotion analysis, as mentioned earlier, combining text, audio, and visual data [36]. Transformers' uniform attention mechanism integrates these data streams, giving robots a holistic view of a situation. A multimodal transformer could take a user's word transcription (for semantic content via an NLP submodel), audio waveform (for vocal tone via an audio transformer), and face image (for expression via a vision transformer) as input and output an emotion classification or appropriate response. Though still in development, such systems aim to give robots human-like perceptual abilities. Recent survey (2024) on transformer-based multimodal emotion recognition models shows new state-of-the-art results on test datasets, confirming their effectiveness [45].

4.3.3. Empathy and personality generation:

Transformers can also give robots personalities, affecting emotional exchange. A language model can change tones by training or prompting it with a persona (a friendly older mentor, a cheerful child, a professional customer service agent). AI can define a social robot's "character" including its emotional expression and sensitivity. An affective dialogue model could be tuned on the Facebook AI "Empathetic Dialogues" corpus to respond with empathy to personal stories. This would make a robot companion better at talking about feelings; a user complaining about a bad day would get a compassionate response [46].

4.3.4. Vision and action generation:

The output of generative models can help robots express emotions beyond text. Generative adversarial networks (GANs) or VAEs are being used to create realistic robot facial expressions or emotionally intoned speech via neural TTS. As a sequence prediction problem, a transformer-based model could choose the robot's next facial expression or gesture frame-by-frame. In response to context, a transformer controlling a robot's torso may generate joint movements for a "sad posture" or a "happy dance". This is younger than NLP but growing in line with the idea of robots that act out emotions realistically [25].

In essence, transformer models and contemporary ML breakthroughs are supercharging the toolkit for emotional HRI. They make the robot's side of the interaction far more flexible and context-aware than older hard-coded systems. A decade ago, a "sympathetic" robot might have had a few canned phrases and expressions. Today, with powerful AI, we can imagine robots that dynamically generate responses to what a user said, possibly even with a touch of creativity or humour. This greatly enhances the illusion of understanding. The fluency and contextual coherence generated by these models reduce the cognitive load required to interpret the interaction, thereby lowering critical scrutiny and facilitating the automatic processing characteristic of cognitive illusion [31,36]. When a robot responds coherently, on topic, and empathetically, it's easy to believe it has intelligence and feelings. Remember that even the most advanced AI relies on pattern recognition and reinforcement of learnt examples, not emotion. If users, especially children or those unaware of AI's limitations, trust the robot like a friend, danger arises. A recent concern highlighted the "empathy gap" in AI chatbots with children: kids may share private feelings with a chatbot, assuming it cares and is trustworthy, which could have complex psychological consequences [47]. Implementing advanced models requires user education and ethical guardrails, such as monitoring AI therapy bots for unsafe advice and ensuring children understand their robot friend's identity.

Having covered the technological advancements, we now turn to applications where human-robot emotional interactions are particularly salient. We will focus on two domains the user is interested in: healthcare and customer service, examining how the illusion of reciprocity is utilized, the benefits observed, and the challenges faced in each.

5. Applications in Healthcare: Robots for Emotional Support and Care

A recent meta-analysis specifically examining pet therapy in elderly populations provides strong empirical support for these interventions. Manikandan et al. [48] conducted a restructured meta-analysis of 11 studies involving 494 participants and reported a pooled effect size of SMD equals negative 0.92 (95% CI: negative 1.28 to negative 0.56, p less than 0.001). The authors note that this large effect corresponds to approximately a 35% to 40% reduction in depression scores on clinical scales. The effectiveness of these interactions is rooted in specific neuroendocrine mechanisms. As highlighted by Magiera et al. [49], interaction with social agents correlates with higher concentrations of oxytocin, associated with bonding and trust, alongside simultaneous lower levels of cortisol, the body's primary stress hormone. Robotic alternatives like Paro are explicitly designed to mimic these biological triggers. By simulating tactile stimulation and contingent feedback, these robotic agents can stimulate oxytocin secretion and lower blood pressure, thereby eliciting physiological relaxation responses similar to those observed in animal-assisted therapy [14].

Healthcare is a major market for emotional social robots. Robots are helping with tasks, providing emotional support, companionship, and therapeutic interactions in eldercare, mental health therapy, and paediatric care. Even though robots have no feelings, patients and carers may find comfort or motivation in them. Examples and implications of emotional HRI in healthcare are discussed below. Key applications and findings in healthcare robotics are summarized in **Table 3**.

Table 3. Applications in Healthcare (Robots for Emotional Support)

Study	Strengths	Key Findings	Limitations
[19]	Effectively reduces stress and anxiety in patients	Patients perceive Paro as affectionate, benefiting those with dementia	Ethical concerns regarding the illusion of care and emotional attachment
[14]	Engage patients with friendly behaviour in hospitals	Reduces anxiety in hospital settings by offering entertainment and companionship	Limited emotional depth – more a social lubricant than a caregiver
[41]	Engages children in educational contexts with empathy-driven interaction	Children improve social skills and emotional recognition through interactions with Tega	Over-reliance could hinder real human-to-human social development

Two elderly patients interact with “Paro,” a therapeutic robot designed as a baby harp seal. Paro responds to touch and sound, exhibiting behaviours like moving its head and purring when petted. This pet-like presence can soothe and engage dementia patients, alleviating loneliness and anxiety [14,19]. Paro is a popular eldercare social robot. Over the past decade, Paro has been shown to reduce stress, agitation, and pain in dementia and chronic illness patients. It mimics the benefits of animal therapy (lowering blood pressure and improving mood) without the risks. Paro is comforting to patients who talk to and cuddle it. They often think Paro likes them or is happy to be with them. This is the illusion of reciprocity: an affectionate object meets the patient's emotional needs for a carer. Paro's success in over 30 countries shows that many carers accept this trade-off for patient well-being [50].

Is it ethical to “trick” dementia patients into bonding with robots like Paro? Advocates say reducing distress is the priority for patients who don't understand the difference and even those who do. Alzheimer's patients can sleep better and feel less anxious if they think a fluffy robot cares about them. Paro often calms inconsolable patients and reduces the need for sedatives, according to clinical staff [47]. Detractors worry that patients may be misled or that robots could be used to reduce human attention. Paro appears to be used as a supplement to human care, not as a replacement. Carers should monitor how patients use Paro and make sure it's in their best interest, not just for staff convenience. Leaving a patient with a robot all day instead of personal interaction is problematic. In conclusion, Paro shows how the illusion of a reciprocal affectionate relationship (patient-pet) can lead to real therapeutic gains, and ethics are generally accepted because the patient's benefit outweighs the abstract concern of deception. The robot can mimic facial expressions, play games, and patiently ask the child questions. The robot may encourage the child ("I'm happy you showed me a smile!"), and the child may personify the robot ("NAO is my friend who plays with me"). Robot buddies improve children's communication and emotional recognition, according to research. Therapists can use children's assumptions about the robot's emotions to motivate them ("Oh, NAO is sad you didn't answer, can you try saying hello back to NAO?"). Since children usually know the robot is a robot (or a machine), the ethical consensus is usually positive. It aids skill development. Therapists use the robot as a tool, not the sole social partner, because a child may become too attached to it or prefer robots to people.

In hospitals, beyond psychological support, robots are being used to improve the patient experience. For example, Pepper has been placed in some hospital lobbies to greet visitors and patients, provide information, and lighten the mood. Pepper's smile and chattiness can ease hospital visits. It may joke or entertain in waiting areas. This is more about customer service, but patients may feel more welcome and can ask simple questions to the robot, freeing up staff. Pepper's role in these situations is social lubrication, not emotional support. Pepper is not a doctor or nurse, so people probably treat it as a helpful gadget with personality. Pepper's design does try to make people feel seen and cheered up in brief interactions, which is emotional exchange (albeit shallow).

5.1 Patient and caregiver acceptance:

Interestingly, studies on robot caregivers often find that familiarity and severity of need influence acceptance [51]. Social robots were more accepted in severe conditions (where help is urgently needed) than in intimate or emotional roles that could replace human contact in one expert survey. Using a robot to remind and companion a dementia patient was acceptable, but using one as a child's playmate or an adult's "romantic partner" was not. It seems ethical to avoid robots in emotionally intimate or high-stakes situations. Healthcare is in the middle and high-stakes for patient well-being, so robots are welcome where they can help fight depression. Since it's a caring domain, we carefully evaluate their use. Carers worry that patients may become attached to a robot and be upset if it breaks or is removed. Managing the end of a human-robot relationship (Patient moves to a facility without Paro) is crucial [52].

5.2 Privacy and data:

Another concern in healthcare HRI is privacy, since emotional robots often collect sensitive data (camera feeds, microphone recordings of a patient's words). All the emotion recognition and adaptive behaviour come from analyzing the patient effectively [53]. Hospitals and developers must ensure compliance with privacy laws (like HIPAA, GDPR) if this data is stored or processed in the cloud [54]. And on a personal level, trust can be eroded if a patient feels the robot is "spying" or not respecting their confidentiality. Transparency about data handling and giving patients or guardians control over what the robot records can mitigate.

In healthcare applications, emotionally capable robots have shown significant promise. Robots provide companionship, support, and emotional and physical health improvements, while respecting patient dignity and rights. With aging populations and caregiver shortages, more robots are expected. Ongoing research and pilot programs (for example, using robots to support mental health interventions like social skills training for schizophrenia, or providing companionship for isolated COVID-19 patients) will continue to test where these tools are most effective. Each success will likely come with the need to carefully manage that human tendency to treat the robot as more than a machine [55].

6. Applications in Customer Service: Empathic Bots and Social Retail Robots

Customer service domains, including retail, hospitality, and online support, leverage human-robot (HAI) emotional interaction to optimize customer satisfaction, user experience (UX), and operational efficiency. Unlike healthcare, where the primary objective is patient well-being, the goal in customer service is pragmatic: positive emotional interactions can foster brand loyalty and increase efficiency. However, the illusion of reciprocity remains a central mechanism, as a responsive agent can significantly improve customer perception even during problem resolution. Key applications and findings in customer service robotics are summarized in **Table 4**.

Table 4. Applications in Customer Service (Empathetic Bots and Social Robots)

Study	Strengths	Key Findings	Limitations
Pepper [56]	Increases engagement in retail settings	Customers enjoy the novelty and personality of Pepper, boosting business interactions.	Lacks deep empathy and mainly serves as a novelty
Hilton Concierge [14]	Offers efficient service with a friendly demeanor	Guests appreciate consistency and politeness in interactions with robot concierges.	Cannot handle complex or sensitive customer needs
Chatbot [57]	Effective in managing customer emotions and satisfaction	Empathy-based chatbots increase customer trust and improve service quality.	Overuse of empathetic language can backfire if not paired with proper solutions

6.1 Embodied Robots: The Social Lubricant Effect

In physical retail and hospitality settings, humanoid robots such as SoftBank's Pepper are deployed as social lubricants to enhance the customer environment. Technically, these robots utilize multimodal emotion recognition systems that integrate data from cameras (facial expressions) and microphones (vocal prosody). As discussed in Section 4.1, these systems often employ Swallowing or Mixing Fusion architectures to accurately classify customer states, such as happiness or confusion.

Pepper, for example, is designed to engage people through friendly appearance and interactive behaviors. It can analyze basic emotional cues, such as recognizing a smile or frown, and adapt its responses accordingly. However, its primary application remains promotional or for simple Q&A. In retail stores and malls, Pepper often functions as a greeter or guide, executing entertainment behaviors such as dancing or telling jokes to create a pleasant atmosphere. This novelty aspect serves to attract customers and improve mood, rather than providing deep empathetic resolution [56].

Similarly, in the hospitality sector, robots are employed as concierges or butlers. Hilton's test of the robot concierge "Connie" and the use of humanoid robots in Japanese hotels illustrate this trend. These systems are programmed with polite, soothing voices to handle common traveler questions with unwavering patience and courtesy[14]. While they cannot handle complex or sensitive issues requiring human judgment, they maintain a consistent emotional state, providing a reliable baseline of service that an exhausted human clerk might struggle to match.

6.2 Chatbots, LLMs, and the Risk of "Insincere Empathy

In the realm of online customer service, emotional AI is operationalized through Transformer-based Natural Language Processing (NLP) models. Rather than relying on keyword matching, modern chatbots utilize the self-attention mechanism to parse the semantic subtext of customer queries. This allows systems to perform sentiment analysis, categorizing messages as "angry" or "sad" to prioritize urgent issues and trigger empathetic response templates [58].

However, the illusion of empathy in these systems is fragile. Research indicates that if an empathetic response is not backed by a substantive solution, the illusion breaks, leading to user distrust. Lee et al. (2025) demonstrated that offering an angry customer an empathetic apology alongside a low-value coupon had a worse outcome than offering the coupon with a neutral message [25]. This phenomenon suggests that users perceive disingenuous emotional simulation as manipulation. Consequently, designers must calibrate the level of simulated empathy to match the severity of the customer's issue to avoid the "backfire" effect of perceived insincerity.

6.3 The “Judgement-Free” Zone: Venting to Machines

Interestingly, the asymmetry of human-robot interaction offers a unique psychological benefit in customer service. Because customers know these service robots are not sentient, they often engage with them more freely than they would with a human. Harvard studies have found that customers exhibit higher self-disclosure and emotion when talking to bots, likely because they do not fear judgment or adverse reaction, provided the bot responds empathetically and helpfully [59]. This suggests that the illusion need not be perfect; users may partially realize it is an illusion but still engage in the dynamic because it serves their emotional needs, creating a safe space for venting frustration.

6.4 Future Directions and Ethical Constraints

Looking forward, the integration of emotional AI in customer service presents a dual-edged prospect. While seamless AI agents could drastically reduce costs while maintaining satisfaction, there is a risk that customers will feel companies are "fobbing them off" to unfeeling machines [60]. Transparency remains a critical constraint; as bots become more human-like, users may struggle to distinguish them from human agents [61].

Ultimately, emotional interaction with robots/AI improves customer service by offering instant, judgment-free, and context-aware support. While customers often realize the interaction is simulated, a well-designed empathetic interface, whether an embodied robot like Pepper or a text-based chatbot, can still deliver high satisfaction. The success of this domain demonstrates that pragmatic emotional AI can effectively scale social dynamics that were previously limited to human agents [62].

7. Challenges and Future Directions

The progression of emotionally intelligent robots and AI agents presents both exciting opportunities and serious challenges. As we deepen the illusion of reciprocity with ever more sophisticated technology, we must address the ethical, social, and technical issues that arise [63]. Key challenges and future considerations are summarized in **Table 5**.

Table 5. Challenges and Ethical Implications in HRI

Study	Strengths	Key Findings	Limitations
[14]	Ethical guidelines for robot interaction	Highlights the ethical dilemma of robots "pretending" to be empathetic	Lacks concrete solutions for preventing emotional manipulation

Continuation Table:

Study	Strengths	Key Findings	Limitations
[26]	Insight into the societal implications of HRI	Emotional robots could alleviate loneliness, but may cause dependency	Ethical dilemma of emotional reliance on robots without true empathy
[22]	Focus on transparency in HRI	Stresses the importance of maintaining transparency regarding robot capabilities	Ethical dilemma of emotional reliance on robots without true empathy

Here we outline key challenges and future considerations in detail, categorized to align with the interdisciplinary framework of this review.

7.1 Authenticity and overcoming the “Fake” Barrier:

Regardless of technical advancement, a robot's empathy remains simulated. Users may eventually encounter what we term an "uncanny valley of empathy", the point where they recognize that the robot’s responses, though perfectly polite or caring, lack the authenticity of human empathy. This realization can result in disappointment or a breakdown of trust [64].

To mitigate this, researchers suggest adopting a strategy of "transparent empathy." This involves AI that supports and emotionally aligns with the user while explicitly acknowledging its artificial nature. For example, a robot stating, "I wish I could offer you a hug, but I am a machine designed to assist," may paradoxically increase trust by managing expectations. Designers must balance the goal of creating beneficial relationships with the necessity of avoiding long-term deception.

7.2 Emotional Dependency and Social Impact

A significant societal concern is whether individuals will prefer robot companionship or AI advice over human interaction. If elderly individuals rely primarily on robots for conversation, they may become more isolated from family and community, as these human connections atrophy. Attachment to a non-reciprocal agent may leave psychological gaps or vulnerabilities.

Similarly, children raised with AI buddies may struggle with the complexities of human relationships, which require mutual accommodation unlike compliant robots. Experts recommend that social robots be designed to complement, rather than replace, human interaction, for instance, by prompting a child to invite a friend to play together rather than playing alone with the robot [65]. As robots become standard, societal norms may need to evolve to teach children to treat robots with kindness while understanding their limitations.

7.3 Privacy and Ethical Use of Data

Emotional AI systems rely on multimodal inputs such as facial expressions, vocal tone, and potentially biometric data. This raises significant privacy concerns. Users may not want companies to record their emotions or store reaction videos. There is a risk that psychological data could be misused to manipulate consumer behavior.

Ethical guidelines and regulations will likely be required to ensure emotional data is not exploited beyond the immediate purpose of assisting the user. In healthcare, such data may be considered medical information and protected accordingly under laws like HIPAA or GDPR [54]. An emerging concept is "affective rights", the idea that a person’s emotional signals and profiles belong to them and should be safeguarded.

7.4 Bias and Cultural Differences

Cultures and individuals express emotions differently. AI models are only as good as the data they are trained on. A system trained predominantly on English or Western facial expressions may misinterpret a user from a different culture. For instance, facial expression software has been shown to label neutral faces of certain ethnic groups as angry more frequently due to bias in training datasets [66].

This presents a technical and ethical risk: a robot might misinterpret a customer’s natural expression as anger and apologize unnecessarily, or worse, adjust its behavior inappropriately. Future systems must be trained on diverse populations and incorporate context-dependent cultural interpretation frameworks to prevent emotional AI bias and misinterpretation.

7.5 Safety and Emotional Contagion

Teammates can share and escalate emotions. A robot that reflects a user's anger back at them, for instance, responding sternly to match assertiveness, could escalate a situation rather than de-escalate it. For emotional safety, robots must be programmed to calm humans, not aggravate them. Calm, consistent verbal responses are less likely to provoke conflict than a reactive, harsh tone [67].

Furthermore, the concept of robot respect is vital. While a human user may lose patience, the robot should remain infinitely patient. If robots were programmed to simulate frustration, conflicts could arise. Future robots should generally avoid expressing negative emotions toward users, except in narrow, controlled simulations (such as a grumpy robot in a game).

7.6 Legal and Policy Frameworks

Legal ambiguities arise as emotionally interactive robots proliferate. Who is liable if a robot causes emotional harm, for example, if a therapy bot gives advice that worsens a patient's depression, or if a child is traumatized when a companion robot is removed? Can "emotional neglect" by a caregiving robot be considered malpractice?

Robots used in sensitive fields like mental healthcare may require strict standards and certifications. Organizations like IEEE are developing ethical standards for AI, and some governments are considering robot-specific regulations. Policy issues include requirements for transparency (identifying the agent as AI) and informed consent for using a user’s emotional data. The coming years will likely see these frameworks solidified [68].

7.7 Human-in-the-Loop and Hybrid Models

A promising approach to managing the illusion of reciprocity, which is the mistaken belief that AI systems genuinely understand or care about human emotions, is the integration of Human-in-the-Loop (HITL) systems. In such models, AI or robotic agents handle routine interactions, while complex or emotionally charged situations are escalated to human supervisors [69]. For instance, an AI-powered customer service bot can detect rising user frustration through real-time sentiment analysis and seamlessly transfer the conversation to a human agent. Similarly, in therapeutic contexts, clinicians may monitor a companion robot’s interactions via real-time dashboards and intervene when signs of distress or ethical concern arise.

This hybrid architecture leverages the scalability and efficiency of AI while preserving the empathy, contextual ethical reasoning, and adaptability of human caregivers. Looking ahead, future systems could extend this paradigm to “AI overseeing AI,” wherein a secondary autonomous module continuously analyzes interaction patterns for indicators of user distress or ethical risk, alerting human supervisors only when necessary, thereby optimizing both safety and resource efficiency

8. Conclusion

The pursuit of emotionally intelligent robots transcends a technical endeavor; it represents a profoundly human undertaking that demands wisdom, ethical foresight, and interdisciplinary collaboration. As reviewed in this paper, and as synthesized in Figure 1, the emotional bond between humans and robots is fundamentally an illusion of reciprocity. It is an asymmetrical exchange where humans experience genuine affect, and machines simulate empathy through advanced multimodal fusion and generative architectures. We have demonstrated that this illusion is not merely a philosophical curiosity but a psychological phenomenon grounded in enactive cognition, where synchronization of auditory, visual, and tactile cues creates a perception of social agency.

Despite this asymmetry, the utility of this illusion is empirically supported. As highlighted in our review of healthcare applications, meta-analytical evidence confirms that simulated emotional support can yield significant physiological and psychological benefits, such as reduced cortisol levels and improved depressive symptoms. However, these benefits are not without risk. The same Transformer-based NLP models and Reinforcement Learning algorithms that enable adaptive, empathetic responses also raise the specter of dependency, privacy invasion, and algorithmic bias.

The future of HRI, therefore, lies not in making robots indistinguishable from humans, but in managing the ethical boundaries of the illusion. Our analysis suggests that pure automation is insufficient. Instead, Human-in-the-Loop (HITL) and hybrid models represent the most viable path forward. By integrating human oversight for emotionally charged or sensitive contexts, we can harness the scalability and consistency of AI while preserving the moral judgment and authenticity of human connection. Success will be defined by our ability to design systems that amplify human well-being through this simulated exchange, without eroding the fabric of authentic human relationships.

Author contributions: Conceptualization, WPP, TAA and ADC; investigation, WPP, TAA and KDV; writing—original draft, WPP, TAA and KDV; writing—review and editing, TAA, ADC, SZZ, EYZ.and KDV; supervision, ADC; project administration, TAA. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors extend their sincere gratitude to Nanjing University of Information Science and Technology and JMRAI Co., Ltd. Jiangyin for their institutional support and stimulating scholarly environments, which were instrumental to this work.

Conflict of Interest: The authors declare no conflict of interest.

Declaration of AI-Assisted Writing: During the revision of this manuscript, the authors utilized Qwen3.6 (an AI language model developed by Alibaba Cloud) for language polishing, grammar checking, and improving sentence fluency. All AI-generated suggestions were reviewed, edited, and approved by the authors to ensure accuracy and alignment with the scientific content. The core ideas, analysis, and final intellectual content remain the sole responsibility of the authors.

References

1. de Graaf MMA. An Ethical Evaluation of Human-Robot Relationships. International Journal of Social Robotics. 2016, 8(4): 589–598. https://doi.org/10.1007/s12369-016-0368-5.
2. Sahu OP, Karmakar M. Disposable culture, posthuman affect, and artificial human in Kazuo Ishiguro's Klara and the Sun (2021). AI & Society. 2024, 39(3): 1349–1357. https://doi.org/10.1007/s00146-022-01600-1.
3. Prescott TJ, Robillard JM. Are friends electric? The benefits and risks of human-robot relationships. iScience. 2021, 24(1): 101993. https://doi.org/10.1016/j.isci.2020.101993.
4. Pouw W, Trujillo JP, Dixon JA. The quantification of gesture-speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. Behavior Research Methods. 2020, 52(2): 723–740. https://doi.org/10.3758/s13428-019-01271-9.
5. Wiese E, Metta G, Wykowska A. Robots As Intentional Agents: Using Neuroscientific Methods to Make Robots Appear More Social. Frontiers in Psychology. 2017, 8: 1663. https://doi.org/10.3389/fpsyg.2017.01663.
6. Cross ES, Hortensius R, Wykowska A. From social brains to social robots: applying neurocognitive insights to human–robot interaction. Philosophical Transactions of the Royal Society B: Biological Sciences. 2019, 374(1771): 20180024. https://doi.org/10.1098/rstb.2018.0024.
7. Seifert J, Friedrich O, Schleidgen S. Imitating the Human. New Human-Machine Interactions in Social Robots. Nanoethics. 2022, 16(2): 181–192. https://doi.org/10.1007/s11569-022-00418-x.
8. Konijn EA, Hoorn JF. Differential Facial Articulacy in Robots and Humans Elicit Different Levels of Responsiveness, Empathy, and Projected Feelings. Robotics. 2020, 9(4): 92. https://doi.org/10.3390/robotics9040092.
9. Xue T, Wang W, Ma J, et al. Progress and Prospects of Multimodal Fusion Methods in Physical Human-Robot Interaction: A Review. IEEE Sensors Journal. 2020, 20(18): 10355–10370. https://doi.org/10.1109/JSEN.2020.2995271.
10. Yang Y, Cho HK, Kim MY. A Study on Factors Affecting the Continuance Usage Intention of Social Robots with Episodic Memory: A Stimulus-Organism-Response Perspective. Applied Sciences. 2025, 15(10): 5334. https://doi.org/10.3390/app15105334.
11. Stock-Homburg R. Survey of Emotions in Human-Robot Interactions: Perspectives from Robotic Psychology on 20 Years of Research. International Journal of Social Robotics. 2022, 14(2): 389–411. https://doi.org/10.1007/s12369-021-00778-6.
12. Baek F, Kim D, Lee G, et al. Effect of human emotional responses on human-robot team pty in construction. Automation in Construction. 2024, 166: 105608. https://doi.org/10.1016/j.autcon.2024.105608.
13. Hornecker E, Bischof A, Graf P, et al. The Interactive Enactment of Care Technologies and its Implications for Human-Robot-Interaction in Care. Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. New York, NY, USA: ACM. 2020: 1–11. https://doi.org/10.1145/3419249.3420103.
14. Korn O, editor. Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction. Cham: Springer International Publishing. 2019. https://doi.org/10.1007/978-3-030-17107-0.
15. Hsieh TY, Chaudhury B, Cross ES. Human-Robot Cooperation in Prisoner Dilemma Games. Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. New York, NY, USA: ACM. 2020: 257–259. https://doi.org/10.1145/3371382.3378309.
16. Correia F, Mascarenhas SF, Gomes S, et al. Exploring Prosociality in Human–Robot Teams. 2019 14th ACM/IEEE International Conference on Human–Robot Interaction (HRI). IEEE. 2019: 143–151. https://doi.org/10.1109/HRI.2019.8673299.
17. Sharkey A, Sharkey N. We need to talk about deception in social robotics!. Ethics and Information Technology. 2021, 23(3): 309–316. https://doi.org/10.1007/s10676-020-09573-9.
18. Rosero A, Dula E, Kelly H, et al. Human perceptions of social robot deception behaviors: an exploratory analysis. Frontiers in Robotics and AI. 2024, 11: 1409712. https://doi.org/10.3389/frobt.2024.1409712.
19. Fron C, Korn O. A Short History of the Perception of Robots and Automata from Antiquity to Modern Times. 2019: 1–12. https://doi.org/10.1007/978-3-030-17107-0_1.
20. Corti L, Di Stefano N, Bertolaso M. Artificial Emotions: Toward a Human-Centric Ethics. International Journal of Social Robotics. 2023, 15(12): 2039–2053. https://doi.org/10.1007/s12369-022-00890-1.
21. van Straten CL, Peter J, Kühne R, et al. Transparency about a Robot's Lack of Human Psychological Capacities. ACM Transactions on Human-Robot Interaction. 2020, 9(2): 1–22. https://doi.org/10.1145/3365668.

22. Yu L, Ding S. Ethics and risks between human and robotic interaction. Interaction Studies. 2019, 20(1): 134–147. https://doi.org/10.1075/is.18009.yu.
23. Ekbia HR. Alone Together: Why We Expect More from Technology and Less from Each Other by Sherry Turkle. Journal of the American Society for Information Science and Technology. 2012, 63(9): 1897–1898. https://doi.org/10.1002/asi.22658.
24. Pagallo U. Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots. Information. 2018, 9(9): 230. https://doi.org/10.3390/info9090230.
25. Lee J. When bots' empathic expressions backfire: Exploring the negative moderation between chatbot empathic expressions and monetary compensation for angry customers. Electronic Markets. 2025, 35(1): 65. https://doi.org/10.1007/s12525-025-00814-7.
26. Naneva S, Sarda Gou M, Webb TL, et al. A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots. International Journal of Social Robotics. 2020, 12(6): 1179–1201. https://doi.org/10.1007/s12369-020-00659-4.
27. Anggraini ID, Sarah NN. Human-Robot Relationships in Millennial CGI Narratives: Ethical, Emotional, and Social Dimension. Scope: Journal of English Language Teaching. 2024, 9(1): 236. https://doi.org/10.30998/scope.v9i1.24705.
28. Marcos-Pablos S, García-Peñalvo FJ. Emotional Intelligence in Robotics: A Scoping Review. 2022: 66–75. https://doi.org/10.1007/978-3-030-87687-6_7.
29. Akinpelu S, Viriri S, Adegun A. An enhanced speech emotion recognition using vision transformer. Scientific Reports. 2024, 14(1): 13126. https://doi.org/10.1038/s41598-024-63776-4.
30. Vaaras E, Ahlqvist-Björkroth S., Drossos K, et al. Development of a speech emotion recognizer for large-scale child-centered audio recordings from a hospital environment. Speech Communication. 2023, 148: 9–22. https://doi.org/10.1016/j.specom.2023.02.001.
31. Rezapour M. Emotion Detection with Transformers: A Comparative Study. arXiv. 2024, abs/2403.15454.
32. Acheampong FA, Nunoo-Mensah H, Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches. Artificial Intelligence Review. 2021, 54(8): 5789–5829. https://doi.org/10.1007/s10462-021-09958-2.
33. Li M, Yu P, Shen Y. A spatial and temporal transformer-based EEG emotion recognition in VR environment. Frontiers in Human Neuroscience. 2025, 19: 1517273. https://doi.org/10.3389/fnhum.2025.1517273.
34. Qiu L, Zhong L, Li J, et al. SFT-SGAT: A semi-supervised fine-tuning self-supervised graph attention network for emotion recognition and consciousness detection. Neural Networks. 2024, 180: 106643. https://doi.org/10.1016/j.neunet.2024.106643.
35. Spezialetti M, Placidi G, Rossi S. Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. Frontiers in Robotics and AI. 2020, 7: 532279. https://doi.org/10.3389/frobt.2020.532279.
36. Hazmoune S, Bougamouza F. Using transformers for multimodal emotion recognition: Taxonomies and state of the art review. Engineering Applications of Artificial Intelligence. 2024, 133: 108339. https://doi.org/10.1016/j.engappai.2024.108339.
37. Chen L, Su W, Feng Y, et al. Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction. Information Sciences. 2020, 509: 150–163. https://doi.org/10.1016/j.ins.2019.09.005.
38. Akalin N, Loutfi A. Reinforcement Learning Approaches in Social Robotics. Sensors. 2021, 21(4): 1292. https://doi.org/10.3390/s21041292.
39. Bhagat S, Banerjee H, Ho Tse ZT, et al. Deep Reinforcement Learning for Soft, Flexible Robots: Brief Review with Impending Challenges. Robotics. 2019, 8(1): 4. https://doi.org/10.3390/robotics8010004.
40. Maroto-Gómez M, Castro-González Á, Castillo JC, et al. A Bio-inspired Motivational Decision Making System for Social Robots Based on the Perception of the User. Sensors. 2018, 18(8): 2691. https://doi.org/10.3390/s18082691.
41. Park HW, Grover I, Spaulding S, et al. A Model-Free Affective Reinforcement Learning Approach to Personalization of an Autonomous Social Robot Companion for Early Literacy Education. Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(1): 687–694. https://doi.org/10.1609/aaai.v33i01.3301687.
42. Holzinger A, Plass M, Holzinger K, et al. Towards interactive Machine Learning (iML): Applying Ant Colony Algorithms to Solve the Traveling Salesman Problem with the Human-in-the-Loop Approach. 2016: 81–95. https://doi.org/10.1007/978-3-319-45507-5_6.
43. Cuayáhuitl H. A data-efficient deep learning approach for deployable multimodal social robots. Neurocomputing. 2020, 396: 587–598. https://doi.org/10.1016/j.neucom.2018.09.104.
44. Chin H, Song H, Baek G, et al. The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study. Journal of Medical Internet Research. 2023, 25: e51712. https://doi.org/10.2196/51712.

45. Borghi M, Mariani MM. The role of emotions in the consumer meaning-making of interactions with social robots. Technological Forecasting and Social Change. 2022, 182: 121844. https://doi.org/10.1016/j.techfore.2022.121844.
46. Kumar A, Sharma K, Sharma A. MEmoR: A Multimodal Emotion Recognition using affective biomarkers for smart prediction of emotional health for people analytics in smart industries. Image and Vision Computing. 2022, 123: 104483. https://doi.org/10.1016/j.imavis.2022.104483.
47. Kurian N. 'No, Alexa, no!': designing child-safe AI and protecting children from the risks of the 'empathy gap' in large language models. Learning, Media and Technology. 2025, 50(4): 621–634. https://doi.org/10.1080/17439884.2024.2367052.
48. Manikandan S, Reshma SR, Raja A, et al. Effectiveness of Pet Therapy in Reducing Depression Among the Elderly: A Restructured Meta-Analysis. Journal of Applied Bioanalysis. n.d., 11.
49. Magiera M, Czwałga P, Sikora M, et al. Animal-Assisted Interventions in Pediatric and Geriatric Care: A Review of Benefits, Mechanisms, and Limitations. Quality in Sport. 2026, 49: 67578. https://doi.org/10.12775/QS.2026.49.67578.
50. Tsiakas K, Karkaletsis V, Makedon F. A Taxonomy in Robot-Assisted Training: Current Trends, Needs and Challenges. Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference. 2018. https://doi.org/10.1145/3197768.3197787.
51. Bharatharaj J, Huang L, Al-Jumaily A, et al. Investigating the Effects of Robot-Assisted Therapy among Children with Autism Spectrum Disorder using Bio-markers. IOP Conference Series: Materials Science and Engineering. 2017, 234: 012017. https://doi.org/10.1088/1757-899X/234/1/012017.
52. Lasota PA, Fong T, Shah JA. A Survey of Methods for Safe Human-Robot Interaction. Foundations and Trends® in Robotics. 2017, 5(4): 261–349. https://doi.org/10.1561/2300000052.
53. Laban G, Cross ES. Sharing Our Emotions With Robots: Why Do We Do It and How Does It Make Us Feel?. IEEE Transactions on Affective Computing. 2026, 17: 1237–1252. https://doi.org/10.1109/TAFFC.2024.3470984.
54. Oruma SO, Ayele YZ, Sechi F, et al. Security Aspects of Social Robots in Public Spaces: A Systematic Mapping Study. Sensors. 2023, 23(19): 8056. https://doi.org/10.3390/s23198056.
55. Obaigbena A, Lottu OA, Ugwuanyi ED, et al. AI and human–robot interaction: A review of recent advances and challenges. GSC Advanced Research and Reviews. 2024, 18(2): 321–330. https://doi.org/10.30574/gscarr.2024.18.2.0070.
56. Xie Y, Zhu K, Zhou P, et al. How does anthropomorphism improve human-AI interaction satisfaction: a dual-path model. Computers in Human Behavior. 2023, 148: 107878. https://doi.org/10.1016/j.chb.2023.107878.
57. Shao C, Kim Y, Xu LZ. From AI Chatbot to Brand Support: An Exploration of Perceived Empathy and Ethics in Shaping Trust and Word of Mouth. Communication Reports. 2026, 39: 69–81. https://doi.org/10.1080/08934215.2025.2519253.
58. Cheng X, Zhang X, Cohen J, et al. Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. Information Processing & Management. 2022, 59(3): 102940. https://doi.org/10.1016/j.ipm.2022.102940.
59. Sahut JM, Laroche M. Using artificial intelligence (AI) to enhance customer experience and to develop strategic marketing: An integrative synthesis. Computers in Human Behavior. 2025, 170: 108684. https://doi.org/10.1016/j.chb.2025.108684.
60. Sidlauskiene J, Joye Y, Auruskeviciene V. AI–based chatbots in conversational commerce and their effects on product and price perceptions. Electronic Markets. 2023, 33(1): 24. https://doi.org/10.1007/s12525-023-00633-8.
61. Pentina I, Hancock T, Xie T. Exploring relationship development with social chatbots: A mixed-method study of replika. Computers in Human Behavior. 2023, 140: 107600. https://doi.org/10.1016/j.chb.2022.107600.
62. Liu X, He X, Wang M, et al. What influences patients' continuance intention to use AI-powered service robots at hospitals? The role of individual characteristics. Technology in Society. 2022, 70: 101996. https://doi.org/10.1016/j.techsoc.2022.101996.
63. Shen JJ, John A St., Alghowinem S, et al. Social Robots as Social Proxies for Fostering Connection and Empathy Towards Humanity. 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2025: 989–999. https://doi.org/10.1109/HRI61500.2025.10973992.
64. Fartook O, McKendrick Z, Oron-Gilad T, et al. Enhancing emotional support in human-robot interaction: Implementing emotion regulation mechanisms in a personal drone. Computers in Human Behavior: Artificial Humans. 2025, 4: 100146. https://doi.org/10.1016/j.chbah.2025.100146.
65. Fang J, Zhou W, Xiong L, et al. The Role of Social Robots in Alleviating Anxiety and Enhancing Mental Well–Being. International Journal of Mental Health and Addiction. 2026, 24(3): 1587–1620. https://doi.org/10.1007/s11469-025-01465-7.
66. Li C, Xing W, Song Y, et al. RICE AlgebraBot: Lessons learned from designing and developing responsible conversational

AI using induction, concretization, and exemplification to support algebra learning. Computers and Education: Artificial Intelligence. 2025, 8: 100338. https://doi.org/10.1016/j.caeai.2024.100338.

67. Seppala E, Rossomando T, Doty JR. Social Connection and Compassion: Important Predictors of Health and Well-Being. Social Research: An International Quarterly. 2013, 80(2): 411–430. https://doi.org/10.1353/sor.2013.0027.
68. Etemad-Sajadi R, Soussan A, Schöpfer T. How Ethical Issues Raised by Human–Robot Interaction can Impact the Intention to use the Robot?. International Journal of Social Robotics. 2022, 14(4): 1103–1115. https://doi.org/10.1007/s12369-021-00857-8.
69. Aradhyula G. Human-in-the-Loop Machine Learning Systems. World Journal of Advanced Engineering Technology and Sciences. 2024, 11(1): 514–521. https://doi.org/10.30574/wjaets.2024.11.1.0012.